A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis

Buch, Stephan ; Stickel, Felix ; Trépo, Eric ; Way, Michael ; Herrmann, Alexander ; Nischalke, Hans Dieter ; Brosch, Mario ; Rosendahl, Jonas ; Berg, Thomas ; Ridinger, Monika ; Rietschel, Marcella ; McQuillin, Andrew ; Frank, Josef ; Kiefer, Falk ; Schreiber, Stefan ; Lieb, Wolfgang ; Soyka, Michael ; Semmo, Nasser ; Aigner, Elmar ; Datz, Christian ; Schmelz, Renate ; Brückner, Stefan ; Zeissig, Sebastian ; Stephan, Anna-Magdalena ; Wodarz, Norbert ; Devière, Jacques ; Clumeck, Nicolas ; Sarrazin, Christoph ; Lammert, Frank ; Gustot, Thierry ; Deltenre, Pierre ; Völzke, Henry ; Lerch, Markus M ; Mayerle, Julia ; Eyer, Florian ; Schafmayer, Clemens ; Cichon, Sven ; Nöthen, Markus M ; Nothnagel, Michael ; Ellinghaus, David ; Huse, Klaus ; Franke, Andre ; Zopf, Steffen ; Hellerbrand, Claus ; Moreno, Christophe ; Franchimont, Denis ; Morgan, Marsha Y ; Hampe, Jochen

Abstract: Alcohol misuse is the leading cause of cirrhosis and the second most common indication for liver transplantation in the Western world. We performed a genome-wide association study for alcohol-related cirrhosis in individuals of European descent (712 cases and 1,426 controls) with subsequent validation in two independent European cohorts (1,148 cases and 922 controls). We identified variants in the MBOAT7 ($P = 1.03 \times 10(-9)$) and TM6SF2 ($P = 7.89 \times 10(-10)$) genes as new risk loci and confirmed rs738409 in PNPLA3 as an important risk locus for alcohol-related cirrhosis ($P = 1.54 \times 10(-48)$) at a genome-wide level of significance. These three loci have a role in lipid processing, suggesting that lipid turnover is important in the pathogenesis of alcohol-related cirrhosis.

DOI: https://doi.org/10.1038/ng.3417

A TWO-STAGE GENOME-WIDE ASSOCIATION STUDY CONFIRMS PNPLA3 AND IDENTIFIES TM6SF2 AND MBOAT7 AS RISK LOCI FOR ALCOHOL-RELATED CIRRHOSIS

Stephan Buch[1,34], Felix Stickel[2,34], Eric Trépo[3,4,34], Michael Way[5,6], Alexander Herrmann[1], Hans Dieter Nischalke[7], Mario Brosch[1], Jonas Rosendahl[8], Thomas Berg[8], Monika Ridinger[9,33], Marcella Rietschel[10], Andrew McQuillin[6], Josef Frank[10], Falk Kiefer[10], Stefan Schreiber[11], Wolfgang Lieb[12], Michael Soyka[13], Nasser Semmo[14], Elmar Aigner[15], Christian Datz[15], Renate Schmelz[1], Stefan Brückner[1], Sebastian Zeissig[1], Anna-Magdalena Stephan[1], Norbert Wodarz[9], Jacques Devière[3,4], Nicolas Clumeck[16], Christoph Sarrazin[17], Frank Lammert[18], Thierry Gustot[3,4], Pierre Deltenre[3,19,20], Henry Völzke[21], Markus M. Lerch[22], Julia Mayerle[22], Florian Eyer[23], Clemens Schafmayer[24], Sven Cichon[25], Markus M. Nöthen[26,27], Michael Nothnagel[28], David Ellinghaus[29], Klaus Huse[30], Andre Franke[29], Steffen Zopf[31], Claus Hellerbrand[32], Christophe Moreno[3,4], Denis Franchimont[3,4,35], Marsha Y. Morgan[5,35], Jochen Hampe[1,35]

[1]Medical Department 1, University Hospital Dresden, TU Dresden, Germany
[2]Department of Gastroenterology and Hepatology, University Hospital of Zurich, Switzerland
[3]Department of Gastroenterology, Hepatopancreatology and Digestive Oncology, CUB Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium
[4]Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium
[5]UCL Institute for Liver & Digestive Health, Division of Medicine, Royal Free Campus, University College London, UK
[6]Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, UK
[7]Department of Internal Medicine I, University of Bonn, Germany
[8]Department of Gastroenterology, University Hospital Leipzig, Germany
[9]Dept. of Psychiatry and Psychotherapy, Univ. Regensburg, Germany
[10]Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany
[11]Department of Internal Medicine I, University Hospital Schleswig-Holstein, Campus Kiel, Germany
[12]Institute of Epidemiology and Biobank POPGEN, Christian-Albrechts-University Kiel, Germany
[13]Psychiatric Hospital, University of Munich, Germany
[14]Department of Visceral Surgery and Medicine, Inselspital Bern, Switzerland
[15]Department of Internal Medicine, Hospital Oberndorf, Teaching Hospital of the Paracelsus Private University of Salzburg, Austria
[16]Centre Hospitalier Le Domaine, Université Libre de Bruxelles, Braine-L'Alleud, Belgium
[17]Department of Gastroenterology, University Hospital Frankfurt, Frankfurt/Main, Germany
[18]Department of Medicine II, University Hospital Homburg, Homburg/Saar, Germany
[19]Service d'Hépato-Gastroentérologie, Hôpital de Jolimont, Haine-Saint-Paul, Belgium
[20]Division of Gastroenterology and Hepatology, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Switzerland
[21]Institute for Community Medicine, University Medicine Greifswald, Germany
[22]Department of Internal Medicine A, University Medicine Greifswald, Germany
[23]Department of Clinical Toxicology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[24]Department of General and Thoracic Surgery, University Hospital Schleswig-Holstein, Campus Kiel, Germany
[25]Department of Biomedicine, University Hospital Basel, Switzerland
[26]Institute of Human Genetics, University of Bonn, Bonn, Germany

[27]Department of Genomics, Life and Brain Center, University of Bonn, Germany
[28]Cologne Center for Genomics, University of Cologne, Germany
[29]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
[30]Fritz-Lipmann Institute of Age Research (FLI), Jena, Germany
[31]Department of Internal Medicine 1, University Hospital Erlangen, Germany
[32]Department of Internal Medicine 1, University Hospital Regensburg, Germany
[33]Psychiatric Health Care Aargau, PDAG, Switzerland

[34]*S.B., F.S. and E.T. contributed equally to the study and the manuscript*
[35]*D.F., M.Y,M. and J.H. contributed equally to the study and the manuscript and assume equal responsibility for senior authorship*

To whom correspondence should be addressed:

Felix Stickel, MD
Department of Gastroenterology and Hepatology
University Hospital of Zurich
Rämistrasse 100, CH-8091 Zurich / Switzerland
Phone +41-31-335 7816, Fax +41-31-335 3519
Email: Felix.stickel@ikp.unibe.ch or felix.stickel@usz.ch

LETTER

Alcohol misuse is the leading cause of cirrhosis and the second most common indication for liver transplantation in the Western world[1–3]. We performed the first genome-wide association study for alcohol-related cirrhosis in individuals of European descent (712 cases and 1,426 controls) with subsequent validation in two independent European cohorts (1,148 cases *vs.* 2,315 controls). Variations in the *MBOAT7* ($P = 1.03 \times 10^{-09}$) and *TM6SF2* ($P = 7.89 \times 10^{-10}$) genes were identified as novel risk loci, while rs738409 in *PNPLA3* was confirmed as an important risk locus for alcohol-related cirrhosis ($P = 1.54 \times 10^{-48}$) at a genome-wide level of significance. These three loci play a role in lipid processing suggesting that lipid turnover plays an important role in the pathogenesis of alcohol-related cirrhosis.

Hepatic steatosis develops in most heavy alcohol users but more significant liver injury only develops with persistent alcohol misuse over time; necroinflammation and progressive fibrosis will develop in 10-35 % of individuals while cirrhosis will develop in only 10-15 %[4]. Thus, only a minority of long-term harmful drinkers will develop cirrhosis. Also, twin studies are pointing to a significant genetic component to disease susceptibility[5,6].

An early GWAS in non-alcohol-related liver disease (NAFLD) identified the variant, rs738409, (148M) in patatin-like phospholipase domain-containing 3 (*PNPLA3*) as a predictor for hepatic fat content[7], and a risk factor for disease progression. This same variant has also been found to be strongly associated with the risk of developing alcohol-related cirrhosis[8,9]. The functional exploration of PNPLA3 has greatly contributed to our understanding of metabolic liver disease[10,11]. To obtain a more complete assessment of the genetic risk structure and to unravel potential new pathways involved in alcohol-related liver disease progression, we performed a two-stage GWAS comparing patients with alcohol-related cirrhosis (cases) to alcohol misusers without cirrhosis (controls).

The first GWAS included 410 Germans with alcohol-related cirrhosis and 1,119 heavy drinkers without liver disease from a German psychiatric consortium[12]. In the second GWAS, 302 cases and 346 controls from the United Kingdom were genotyped (Table 1). Genotype imputation was performed using IMPUTE2[13] to reference panel 1,000 Genomes Phase 3.

Meta-analysis was restricted to markers present in both data sets (N = 6,770,426). The resulting Manhattan plot is provided in Figure 1. The strongest signal was SNP rs738409 located in the *PNPLA3* gene ($P = 1.17 \times 10^{-28}$), further 102 SNPs of genome-wide significance ($P_{meta} < 5 \times 10^{-8}$) located at this locus (Supplementary Table 1).

In addition, a secondary genome-wide meta-analysis adjusted for age, sex, BMI and type 2 diabetes was performed in 456 cases and 873 controls in whom complete covariate information was available (Supplementary Table 2). Genome wide significance was attained only at the *PNPLA3* locus (Supplementary Table 4). The Manhattan and QQ-plots are provided in Supplementary Figures 2 and 3.

As BMI and blood sugar levels are adversely affected by the presence of cirrhosis *per se* these variables may not act as true disease confounders, in this context. In consequence, adjusting the analysis for these variables may instead correct for the intended phenotype. Thus, the most significant variants of the top 10 loci from the primary, *unadjusted* meta-analysis were carried forward to replication. At stage II the SNPs were validated in independent samples from Germany (N = 1,290) and Belgium (N = 779) (Table 1) by fixed-effect model meta-analysis. For the definition of an independent locus, a window of at least 500 kb of genomic distance was required (Supplementary Table 1). In addition to rs738409 in *PNPLA3,* disease association was replicated for variants: rs626283 in *MBOAT7* and rs10401969 in *SUGP1* (Table 2). In the combined analysis of all stage 1-2 samples, both novel replicating SNPs

attained genome-wide significant evidence of association (*MBOAT7:* rs626283 $P_{combined}$ = 1.03×10^{-9}; OR = 1.35 [1.23-1.49] and *SUGP1:* rs10401969 $P_{combined}$= 7.89×10^{-10}; OR = 1.72 [1.44-2.04]; Table 2) and map to distinct loci.

To assess the validity of the design choice made for the primary analysis, a *post-hoc* analysis adjusted for age, sex, BMI and type 2 diabetes was performed for replicating variants. The three loci remain significant in the adjusted analysis and the odds ratios of the adjusted are similar to the unadjusted analysis (Table 3). The top 10 loci from the adjusted GWAS meta-analysis (Supplementary Table 4) were genotyped in the cohorts outlined in Supplementary Table 2. Only *PNPLA3* rs738409 variant was replicated with a combined adjusted *P*-value of 2.16×10^{-26} (Supplementary Table 3).

The association of alcohol-related cirrhosis with the three validated loci was assessed for each of the three case cohorts (Germans, UK and Belgium) on a population level. The odds ratios for the presence of the risk allele in individuals from the general population were lower in comparison to those of heavy drinkers without liver disease *PNPLA3* (OR_{pop} = 1.66 [1.49 - 1.83] vs. OR_{alc} = 2.21 [1.98 - 2.43]), *TM6SF2* (OR_{pop} = 1.48 [1.24 - 1.72] vs. OR_{alc} = 1.63 [1.47 - 1.80]) and *MBOAT7* (OR_{pop} = 1.19 [1.10 - 1.30] vs. OR_{alc} = 1.35 [1.26 - 1.44]) (Figure 2). Population Attributable Risks (PAR) were calculated against population controls and controls drinking harmfully but free of liver injury (Supplementary Table 5). The PAR% in individuals drinking harmfully ranged from 20.6% to 27.3% for PNPLA3; 2.5 % to 5.2 % for TM6SF2; and 7.4 % to 17.2 % for MBOAT7 (Supplementary Table 5).

The strongest signal at the *MBOAT7* locus was obtained for the imputed variant rs626283 (P = 1.07×10^{-5}; OR = 1.36 [1.19 - 1.57]). The results from the meta-analysis (Figure 3) identified a cluster of variants in high linkage disequilibrium (LD) covering the 5' region of

the neighbouring *TMC4* and *MBOAT7* genes. Fine mapping genotyping and conditional analysis using a total of 13 SNPs from the region confirmed the principal location of the association signal (Supplementary Table 6). The top variant identified through fine mapping, rs641738 is in high LD with rs626283 (r^2 = 0.98). The strongest signal at the *TM6SF2* locus was obtained in the validation cohort for variant rs58542926 (P = 7.34 x 10^{-5}, OR = 1.59 [1.26 - 1.99]) which is in complete LD with *SUGP1* rs10401969 (Table 2) (Supplementary Figure 4).

Cis quantitative trait locus expression (cis-*eQTL*) analyses were performed for *MBOAT7* using the genomic intervals displayed in Figure 3 and publicly available human liver datasets[14]. The *MBOAT7* and *TMC4* transcripts are both expressed in human liver (Supplementary Figure 5). Variant rs641738 showed strong association with expression of *MBOAT7* in the two datasets[14] ($P_{expression} = 1.17 \times 10^{-13}$ and $P_{expression} = 2.23 \times 10^{-3}$). For this variant, significant genotype-specific differential expression for the disease-associated allele was observed in liver tissue from patients with alcohol-related cirrhosis for *MBOAT7* but not for the neighbouring *TMC4* gene (Supplementary Figure 6).

This first genome-wide genetic assessment for alcohol-related cirrhosis confirms the prominent role of *PNPLA3* – a gene identified first in a previous candidate gene association study[9] and confirmed in recent meta-analyses[26,39] – but also identified two novel loci of genome-wide significance which significantly influence risk, *MBOAT7* and *TM6SF2*. Associations persist after adjustment for gender, age, BMI and type 2 diabetes in a subset of patients (64%) in whom complete phenotype information was available. The observed unadjusted odds ratios for all three loci were similar to those corrected for potential confounder variables supporting their relevance as risk loci for the development of alcohol-related cirrhosis. However, if the primary replication measure were a fully adjusted analysis

then *TM6SF2* would have associated with a nominal p-value of 0.020 and so would not have met Bonferroni-corrected significance.

Other liver-related phenotypes including NAFLD[7,15], chronic hepatitis C[16], gallstone disease[17], primary biliary cirrhosis[18], primary sclerosing cholangitis[19] and drug-induced liver injury [20] have been investigated on a genome-wide scale.-The delay in the systematic investigation of alcohol-related cirrhosis is somewhat surprising. It may in part be due to the difficulties encountered in recruiting appropriate controls – namely individuals exposed to sufficient amounts of alcohol but without evidence of significant alcohol-related liver injury. Very few studies have chosen such approach[8,9]. The importance of using long term heavy drinkers as controls is underlined by the lower odds ratios for all three loci when analysed against the general population controls (Figure 2). To recruit appropriate controls exposed to alcohol but without alcohol-related cirrhosis, we established a collaboration of hepatology and psychiatry centres where relevant dataset were available. In the current study the alcohol-related cirrhosis risk loci are clearly distinct from the reported loci for alcohol dependence, consumption, and withdrawal symptoms[12,21–24]. Further, in candidate gene studies of alcohol dehydrogenase 1B *(ADH1B),* aldehyde dehydrogenase 2 *(ALDH2)* and *GABRA2,* no evidence of association with alcohol-related liver disease was found[25] suggesting that the genetic risk for alcohol dependence and alcohol-related injury are distinct.

Genetic variation at *PNPLA3* has been established as a risk factor for alcohol-related cirrhosis in previous studies[8,9] and meta-analyses[26,39]. The functional significance of the non-synonymous variant rs738409 has been explored[10,11,27], and data from the present study underline its importance on a genome-wide level.

One of the other top-hits identified in this study, rs641738, locates in the region 19q13.42 and contains the *TMC4* and *MBOAT7* genes. Analysis of eQTL data sets indicates that the lead variant alters the expression of the *MBOAT7* gene[14], and that this variant is associated with

genotype-specific expression in liver tissue from patients with alcohol-related cirrhosis (Supplementary Figure 6). These data implicate rs641738 as the functional variant influencing the expression of *MBOAT7* and influencing alcohol-related cirrhosis risk. *MBOAT7* encodes an enzyme with lysophosphatidylinositol acyltransferase activity and has been implicated in anti-inflammatory processes through regulating arachidonic acid levels in neutrophils[28]. In rat liver microsomes, MBOAT7 catalyses the transfer of fatty acid between phospholipids and lysophospholipids[29], a potent driver mechanism of hepatic inflammation. A key molecular product of this enzymatic reaction is sensed by GPR55[30], a lysophosphatidylinositol receptor with cannabinoid sensitivity; this therefore, provides a link between *MBOAT7* and the known role of the endocannabinoid system in hepatic extracellular matrix remodelling[31,32]. These functional observations provide a hypothetical mechanism through which this identified variant may modulate inflammation-driven liver fibrogenesis in alcohol-related steatosis.

The final top-hit identified in the present study, rs58542926 is a coding variant (E167K) in *TM6SF2,* a reported locus for NAFLD phenotypes[15,33], and a protective factor against cardiovascular disease[40]. At a functional level TM6SF2 activity is required for very low density lipoprotein (VLDL) secretion, and the identified variant might result in impaired biological function causing hepatic lipid trapping and, thus, steatosis[15]. *PNPLA3* as well as *TM6SF2* contain functional variants which associate with both alcohol-related and obesity-related liver disease, demonstrating a striking similarity in their heritability, which echo the other similarities in their clinical presentation and histopathology.

A previous GWAS in NAFLD identified a significant association between liver histology and a variant in the neurocan (*NCAN*) gene[34], but later validation in two subsequent GWAS demonstrated that the associations linked to variants within the neighboring *TM6SF2* gene rather than *NCAN*[15,33]. A recent candidate gene study from our group found an association

between the *NCAN* rs2228603 variant and hepatocellular carcinoma[41], but not with cirrhosis *per se*, most likely because this is not the true risk variant for cirrhosis.

In conclusion, the present study confirms *PNPLA3* as a risk locus for alcohol-related cirrhosis and identifies *TM6SF2* and *MBOAT7* as additional risk loci. All three genes are functionally annotated as being involved in lipid metabolic processes. It is known that variants in both *PNPLA3* and *TM6SF2* are associated with increases in intrahepatic fat compatible with loss of function mutations leading to lipid trapping within hepatocytes[10]. Thus, it can be hypothesized that these genetic variants confer risk via dysfunctional lipid turnover. Whether the fine mapping studies identified the variants that exert the functional variability of the coded protein crucial for liver phenotypes requires further in depth research. The top hits overlap with those identified as risk factors for NAFLD implying that both conditions share mechanisms of pathogenesis and that the risk genes may be therapeutic targets in both disorders. Also, the variants at the three identified loci may help to define high risk populations for targeted abstinence intervention and hepatic surveillance programmes[35,36].

Data access: The metanalysis results for all imputed variants and further information are available at http://gengastro.med.tu-dresden.de/suppl/alc_cirrhosis/. Individual-level data are available from the authors upon request.

URL section

The software IMPUTE, SNPTEST and META is accessible under url: http://www.stats.ox.ac.uk/~marchini/software/gwas/gwas.html. BEAGLECALL is accessible under url: http://faculty.washington.edu/browning/beaglecall/beaglecall.html.

REFERENCES

1. Kim, W. R., Brown, R. S., Terrault, N. a & El-Serag, H. Burden of liver disease in the United States: summary of a workshop. *Hepatology* **36,** 227–42 (2002).

2. Rehm, J., Samokhvalov, A. V & Shield, K. D. Global burden of alcoholic liver diseases. *J. Hepatol.* **59,** 160–8 (2013).

3. Burra, P. *et al.* Liver transplantation for alcoholic liver disease in Europe: a study from the ELTR (European Liver Transplant Registry). *Am. J. Transplant* **10,** 138–48 (2010).

4. Teli, M. R., Day, C. P., Burt, A. D., Bennett, M. K. & James, O. F. Determinants of progression to cirrhosis or fibrosis in pure alcoholic fatty liver. *Lancet* **346,** 987–90 (1995).

5. Hrubec, Z. & Omenn, G. S. Evidence of genetic predisposition to alcoholic cirrhosis and psychosis: twin concordances for alcoholism and its biological end points by zygosity among male veterans. *Alcohol. Clin. Exp. Res.* **5,** 207–15 (1981).

6. Reed, T., Page, W. F., Viken, R. J. & Christian, J. C. Genetic predisposition to organ-specific endpoints of alcoholism. *Alcohol. Clin. Exp. Res.* **20,** 1528–33 (1996).

7. Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* (2008).

8. Tian, C., Stokowski, R. P., Kershenobich, D., Ballinger, D. G. & Hinds, D. a. Variant in PNPLA3 is associated with alcoholic liver disease. *Nat. Genet.* **42,** 21–3 (2010).

9. Stickel, F. *et al.* Genetic variation in the PNPLA3 gene is associated with alcoholic liver injury in caucasians. *Hepatology* **53,** 86–95 (2011).

10. Pirazzi, C. *et al.* PNPLA3 has retinyl-palmitate lipase activity in human hepatic stellate cells. *Hum. Mol. Genet.* **23,** 4077–4085 (2014).

11. Smagris, E. *et al.* Pnpla3I148M knockin mice accumulate PNPLA3 on lipid droplets and develop hepatic steatosis. *Hepatology* (2014).

12. Bierut, L. J. *et al.* A genome-wide association study of alcohol dependence. *Arch Gen Psychiatry* **66,** 773 (2009).

13. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).

14. Innocenti, F. *et al.* Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* **7,** e1002078 (2011).

15. Kozlitina, J. *et al.* Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **46,** 352–356 (2014).

16. Patin, E. *et al.* Genome-wide association study identifies variants associated with progression of liver fibrosis from HCV infection. *Gastroenterology* **143,** 1244–52.e1–12 (2012).

17. Buch, S. *et al.* A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat. Genet.* **39,** 995–9 (2007).

18. Mells, G. F. *et al.* Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* **43,** 329–32 (2011).

19. Melum, E. *et al.* Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. *Nat. Genet.* **43,** 17–9 (2011).

20. Daly, A. K. *et al.* HLA-B*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nat. Genet.* **41,** 816–9 (2009).

21. Bierut, L. J. *et al.* A genome-wide association study of alcohol dependence. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 5082–7 (2010).

22. Han, S. *et al.* Integrating GWASs and human protein interaction networks identifies a gene subnetwork underlying alcohol dependence. *Am. J. Hum. Genet.* **93,** 1027–34 (2013).

23. Treutlein, J. *et al.* Genome-wide association study of alcohol dependence. *Arch. Gen. Psychiatry* **66,** 773–84 (2009).

24. Edenberg, H. J. & Foroud, T. Genetics and alcoholism. *Nat. Rev. Gastroenterol. Hepatol.* **10,** 487–94 (2013).

25. Zintzaras, E., Stefanidis, I., Santos, M. & Vidal, F. Do alcohol-metabolizing enzyme gene polymorphisms increase the risk of alcoholism and alcoholic liver disease? *Hepatology* **43,** 352–61 (2006).

26. Chamorro, a-J. *et al.* Systematic review with meta-analysis: the I148M variant of patatin-like phospholipase domain-containing 3 gene (PNPLA3) is significantly associated with alcoholic liver cirrhosis. *Aliment. Pharmacol. Ther.* **40,** 571–81 (2014).

27. He, S. *et al.* A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. *J. Biol. Chem.* **285,** 6706–15 (2010).

28. Gijón, M. a, Riekhof, W. R., Zarini, S., Murphy, R. C. & Voelker, D. R. Lysophospholipid acyltransferases and arachidonate recycling in human neutrophils. *J. Biol. Chem.* **283,** 30235–45 (2008).

29. Yamashita, A. *et al.* Reverse reaction of lysophosphatidylinositol acyltransferase. Functional reconstitution of coenzyme A-dependent transacylation system. *J. Biol. Chem.* **278,** 30382–93 (2003).

30. Moreno-Navarrete, J.M. *et al.* The L-α-lysophosphatidylinositol/GPR55 system and its potential role in human obesity. *Diabetes* **61,** 281–91 (2012).

31. Tam, J. *et al.* Endocannabinoids in liver disease. *Hepatology* **53,** 346–55 (2011).

32. Patsenker, E. *et al.* Cannabinoid receptor type I modulates alcohol-induced liver fibrosis. *Mol. Med.* **17,** 1285–94 (2011).

33. Liu, Y.-L. *et al.* TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5,** 4309 (2014).

34. Speliotes, E.K. *et al.* Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet.* **7,** e1001324 (2011).

35. Di Martino, V., Sheppard, F. & Vanlemmens, C. Early liver transplantation for severe alcoholic hepatitis. *N. Engl. J. Med.* **366,** 478–9; author reply 479 (2012).

36. Vanlemmens, C. *et al.* Immediate listing for liver transplantation versus standard care for Child-Pugh stage B alcoholic cirrhosis: a randomized trial. *Ann. Intern. Med.* **150,** 153–61 (2009).

37. Higgins, J.P.T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21,** 1539–58 (2002).

38. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26,** 2336–7 (2010).

39. Salameh H., *et al*. PNPLA3 gene polymorphism is associated with predisposition to and severity of alcoholic liver disease. *Am. J. Gastroenterol.* **110**, 846-856 (2015).

40. Dongiovanni P, et al. Transmembrane 6 superfamily member 2 gene variant disentangles nonalcoholic steatohepatitis from cardiovascular disease. *Hepatology* **61**, 506-514 (2015).

41. Nischalke, H.D. *et al.* A common polymorphism in the NCAN gene is associated with hepatocellular carcinoma in alcoholic liver disease. *J. Hepatol.* **61**, 1073-1079 (2014).

42. Way, M. *et al.* Genetic variants in or near ADH1B and ADH1C affect susceptibility to alcohol dependence in a British and Irish population. *Addict. Biol.* **20**, 594-604 (2015).

43. Trépo, E. *et al.* Marked 25-hydroxyvitamin D deficiency is associated with poor prognosis in patients with alcoholic liver disease. *J. Hepatol.* **59**, 344–350 (2013).

44. Ishak, K.G., Zimmerman, H.J. & Ray, M.B. Alcoholic liver disease: pathologic, pathogenetic and clinical aspects. *Alcohol Clin. Exp. Res.* **15,** 45–66 (1991).

45. Nguyen-Khac, E. *et al.* Assessment of asymptomatic liver fibrosis in alcoholic patients using fibroscan: prospective comparison with seven non-invasive laboratory tests. *Aliment. Pharmacol. Ther.* **28,** 1188–98 (2008).

46. Sinnott, J.A., Kraft P. Artifact due to differential error when cases and controls are imputed from different platforms. *Hum. Genet.* **131**, 111–119 (2012).

47. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–78 (2007).

48. Browning, B. L. & Yu, Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am. J. Hum. Genet.* **85**, 847–861 (2009).

49. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

50. Roshyara, N. R. & Scholz, M. fcGENE: A Versatile Tool for Processing and Transforming SNP Datasets. *PLoS One* **9**, e97589 (2014).

AUTHOR CONTRIBUTIONS

S.B. performed genotyping, meta-analysis, in silico analysis, drafted and revised the manuscript. F.S. conceptualized the study, recruited subjects, wrote and revised the manuscript. E.T. recruited subjects, validated the study, provided replication data, wrote and revised the manuscript. M.W. recruited subjects, performed genotyping, validation study. A. H. performed bioinformatics work. H.D.N. recruited and phenotyped subjects. M.B. performed expression analysis. J.R., T.B. recruited subjects. M.R., M.R., A.M., J.F., F.K. recruited subjects, performed phenotyping and recruitment of alcoholic controls. S.S. gave technical support, critically revised manuscript. W.L. helped on population genetic statistics. M.S. recruited subjects and phenotyped alcoholic controls. N.S., E.A., C.D., R.S., S.B., S.Z., A.S. recruited subjects. N.W. recruited subjects, performed phenotyping of alcoholic controls. J.D., N.C., C.S., F.L., T.G., P.D. recruited and phenotyped subjects. H.V. recruited population cohort. M.L., J.M., F.E., C.S., recruited and phenotyped subjects. S.C., M.N. performed phenotyping and recruitment of alcoholic controls. M.N. supervised and reviewed statistical analysis. D.E. assisted with bioinformatic analyeis, K.H. performed expression analysis. A.F. gave conceptual advice and bioinformatic support. S.Z., C.H., C.M. recruited subjects. D.F., M.Y.M recruited subjects, drafted and critically revised the manuscript. J.H. conceptualized the study and analytical design, drafted and revised the manuscript. All authors critically revised and contributed to the final manuscript.

ACKNOWLEDGEMENTS

This study was supported by the German Ministry of Education and Research through the Virtual Liver Network (to JH), the PopGen 2.0 network biobank (grant 01EY1103) and institutional funds from the medical faculties of the Technical University Dresden and the Christian-Albrechts-University Kiel and the Swiss National Funds (grant 310030_138747 to FS). The Community Medicine Research net of the University of Greifswald, Germany is funded by the Federal Ministry of Education and Research, the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. MML and JM were supported by the Federal Ministry of Education and Research (BMBF GANI-MED 03152061A, BMBF 0314107), the European Union (EU-FP-7: EPC-TM and EU-FP-7-REGPOT-2010-1) and EFRE-State Ministry of Economics MV (V-630-S-150-2012/132/133). SC, MMN and MR were supported by the German Federal Ministry of Education and Research (BMBF) through the Integrated Networks IntegraMent and Sysmed Alcohol under the auspices of the e:Med Programme (grant 01ZX1314A to MMN and SC and grant 01ZX1311A to MMN and MR). MMN is a member of the DFG-funded Excellence-Cluster ImmunoSensation. Research of HDN related to this project was funded by the Deutsche Krebshilfe (107865). AF's and DE's work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (SysInflame grant 01ZX1306A). This project received infrastructure support from the DFG Excellence Cluster No. 306 "Inflammation at Interfaces". Andre Franke receives an endowment professorship by the Foundation for Experimental Medicine (Zuerich, Switzerland). The UK research effort was funded by a PhD studentship award jointly funded by University College London and the Hobson Charity. We thank colleagues from the following centres for obtaining samples from alcohol dependent cases for genotyping: Bexley Substance Misuse Service, South London & Maudsley NHS Trust; The East Hertfordshire

Community Drug Action Team; the Max Glatt Unit, Southall; Renfrew and Inverclyde Alcohol Services, Strathclyde; Newcastle North Tyneside Drug & Alcohol Service, Tyne and Wear; and from the Acute Admissions Unit and the Centre for Hepatology at the Royal Free Hospital, London. We also thank colleagues associated with the NIHR Mental Health Research Network for assistance in the identification of cases, obtaining consent and collection of samples at the following NHS trusts: Sandwell Mental Health and Social Care; Northamptonshire Healthcare; Avon and Wiltshire Mental Health Partnership, Sheffield Health and Social Care; Tees Esk and Wear Valleys; Lincolnshire Partnership; Nottinghamshire Healthcare; Central and North West London; South Staffordshire and Shropshire Healthcare; Coventry and Warwickshire; and Dudley and Walsall Mental Health Partnership. We are grateful to Jit Saini, Kush Ruparelia, Sara Montagnese, Radhika Kandaswamy, Alexandra Jarram, Giorgia Quadri, and Niamh O'Brien for assisting with the collection and processing of the samples and the DNA extraction.

The Belgian research effort was supported by the Belgian Medical Genomics Initiative (BeMGI) funded by the phase VII Interuniversity Attraction Poles (IAP) program of the Belgian Federal Science Policy Office (BELSPO) and the Fund for Scientific Research - FNRS (F.R.S.-FNRS). ET is a Postdoctoral Researcher of the F.R.S.-FNRS and DF is a Research Director of the F.R.S.-FNRS. We are grateful to Dr. Olivier Lemoine, Dr. Delphine Degré, Dr. Arnaud Lemmers and Mohammed Amrani for the identification of cases and controls, obtaining consent and collection of samples at C.U.B. Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium. We also thank Eric Quertinmont at the Laboratory of Experimental Gastroenterology, Université Libre de Bruxelles, Brussels, Belgium for his help with the collection, the processing of the samples and the DNA extraction.

We are also grateful to the Center for Information Services and High Performance Computing (ZIH) at TU Dresden where the computations were performed on a PC-Cluster.

TABLES

Table 1: Overview of the study populations included in the discovery and validation cohorts and the population controls

Variable	**Discovery** **				**Validation** **				**Non-alcoholic population controls**		
	Germany GWAS 1 (n=1,490)		United Kingdom GWAS 2 (n=648)		Germany (n=1,290)		Belgium (n=780)		Germany	United Kingdom	Belgium
	Cases (n=410)	Controls (n=1,080)	Cases (n=302)	Controls (n=346)	Cases (n=529)	Controls (n=761)	Cases (n=619)	Controls (n=161)	(n=736)	(n=763)	(n=658)
Age* (years)	53 (47-61)	42 (36-48)	53 (47-60)	49 (42-56)	54 (47-62)	46 (39-53)	55 (49-61)	47 (41-55)	64 (57-68)	37 (28-50)	51 (38-61)
Gender (% male)	71	100	68	77	72	83	70	69	50	39	44
BMI*+	26.2 (22.8-29.3)	24.8 (22.7-7.5)	24.8 (22.8-6.8)	24.6 (22.8-26.6)	26.0 (23.0-29.2)	24.3 (21.7-27.0)	25.8 (22.7-9.8)	22.8 (20.6-5.7)	25.9 (23.6-8.4)	N/A	25.2 (22.3-8.4)
Diabetes Type II + (%)	24	4.0	0^	0^	18.1	11.3	18.2	2	9.5	N/A	5.4

*All quantitative measures are median (interquartile range); N/A: not available

**Cases and controls were assigned to groups as detailed in the text and Supplementary Methods.

+ Phenotypic information was available for BMI/Type 2 diabetes as follows:

German discovery cohorts: cases 71 % / 97 %; controls 70 % / 74 %; UK discovery cohorts: cases 80 % / 100 %: controls 53 % / 100 %

German replication cohorts: cases 50 % / 76 %; controls 40 % / 64 %; Belgian replication: cases 85 % / 93 %; controls 91 % / 99 %

^ Individuals with type 2 diabetes and/or BMI >28 were excluded from the UK collections *a priori*

Table 2: Association results for lead markers of regions entering the validation stage of the GWAS.

						Discovery			Validation			Combined*	
SNPs	Locus	Chr	SNP id	Ref. allele	Ref. af	Meta *P value*+ Germany/UK	I^2 *+	OR [95% CI]	Meta *P value* Germany/Belgium	I^2	OR [CI 95]	Meta *P value*	OR [95% CI]
SNP 1	*PNPLA3*	22	rs738409	G	0.27	1.17×10^{-28}	60	2.39 [2.05-2.78]	**4.59×10^{-22}**	17	2.03 [1.76-2.35]	**1.54×10^{-48}**	**2.190 [1.97-2.43]**
SNP 2	*TM6SF2***	19	rs10401969	C	0.08	7.81×10^{-7}	0	1.92 [1.48-2.50]	**1.24×10^{-4}**	0	1.57 [1.25-1.97]	**7.89×10^{-10}**	**1.715 [1.44-2.04]**
		19	rs58542926	T	0.08	2.86×10^{-6}	0	1.87 [1.44-2.43]	**7.34×10^{-5}**	0	1.59 [1.26-1.99]	**1.33×10^{-9}**	**1.704 [1.43-2.02]**
SNP 3	*TM4SF20*	2	rs62190923	G	0.21	1.31×10^{-6}	0	0.64 [0.54-0.77]	0.73	0	1.03 [0.87-1.22]	-	-
SNP 4	*INTERGENIC*	8	rs7812374	T	0.58	1.46×10^{-6}	0	0.70 [0.60-0.81]	0.69	69	0.97 [0.85-1.11]	-	-
SNP 5	*DUSP1*	5	rs6556045	A	0.06	2.51×10^{-6}	0	2.11 [1.55-2.87]	0.16	87	1.20 [0.93-1.56]	-	-
SNP 6	*LPHN2*	1	rs6605237	T	0.27	5.43×10^{-6}	0	1.46 [1.24-1.71]	0.80	0	0.98 [0.84-1.14]	-	-
SNP 7	*IL21*	4	rs17886348	T	0.08	7.64×10^{-6}	60	1.79 [1.39-2.31]	0.99	0	1.00 [0.77-1.30]	-	-
SNP 8	*PDE7B*	6	rs7769670	A	0.14	7.84×10^{-6}	0	1.58 [1.29-1.93]	0.37	0	1.09 [0.90-1.33]	-	-
SNP 9	*INTERGENIC*	8	rs7845021	C	0.61	1.02×10^{-5}	0	0.73 [0.64-0.84]	0.19	0	1.09 [0.95-1.26]	-	-
SNP 10	*MBOAT7*	19	rs626283	C	0.44	1.07×10^{-5}	0	1.36 [1.19-1.57]	**2.29×10^{-5}**	0	1.33 [1.17-1.53]	**1.03×10^{-9}**	**1.349 [1.23-1.49]**

* The results of the combined analyses are only provided for variants meeting a Bonferroni corrected p<0.05 at the validation stage.

** The lead SNP rs739846 failed Taqman genotyping for technical reasons, rs10401969 was used as a replacement (r^2=1.0, $P_{meta\ rs739846}$ =7.45×10^{-7})

The functional variant rs58542926 previously reported at the TM6SF2 locus is included [25, 44]

+ Significance derived from a fixed effect meta-analysis; ++ I^2-measure of percentage of between cohort heterogeneity[37].

Abbreviations: SNP-single nucleotide polymorphism SNP-id SNP identification number; Chr: chromosome; Ref allele---reference allele; Ref af---frequency of the reference allele

Table 3: Validation analysis of the replicated loci in the primary analysis

adjusted for sex, age, BMI and type 2 diabetes, by gender

Locus	Chr	SNP id	Unadjusted Odds Ratio* (p-value)	Adjusted Odds Ratio (p-value)	Gender-specific analysis: Odds Ratios			
					Adjusted, men only	Adjusted women only	Unadjusted** men only	Unadjusted ** women only
PNPLA3	22	rs738409	2.03 [1.76-2.35] 4.59×10^{-22}	2.12 [1.73-2.59] 4.05×10^{-13}	2.07[1.63-2.63]	2.69[1.79-4.04]	2.07[1.73-2.47]	2.44[1.79-3.31]
TM6SF2	19	rs10401969	1.57 [1.25-1.97] 1.24×10^{-4}	1.43[1.06-1.94] 0.020	1.42[0.98-2.05]	1.47[0.78-2.78]	1.51[1.14-1.98]	1.67[1.04-2.68]
	19	rs58542926	1.59 [1.26-1.99] 7.34×10^{-5}	1.43[1.05-1.94] 0.022	1.44[0.99-2.09]	1.41[0.76-2.62]	1.53[1.16-2.02]	1.64[1.03-2.61]
MBOAT7	19	rs626283	1.33 [1.17-1.53] 2.29×10^{-5}	1.41[1.17-1.70] 3.81×10^{-4}	1.19[0.95-1.49]	2.21[1.52-3.21]	1.27[1.08-1.49]	1.51[1.16-1.96]
	19	rs641738	1.35 [1.18-1.54] 1.30×10^{-5}	1.43[1.18-1.72] 2.11×10^{-4}	1.22[0.98-1.52]	2.21[1.52-3.22]	1.29[1.10-1.51]	1.52[1.17-1.98]

* Odds ratios provided with 95 % Confidence Intervals in square brackets ** The unadjusted Odds Ratios are provided for comparison.

Abbreviations: Chr: chromosome; SNP-id single nucleotide polymorphism identification number

FIGURE LEGENDS

Figure 1

Genome-wide association meta-analysis of 712 cases with alcohol-related cirrhosis and 1,466 controls. *P* values are shown for SNPs that passed quality control. The genome-wide significance threshold ($P = 5 \times 10^{-8}$) is indicated as a red line. The threshold for replication follow-up ($P= 2.5 \times 10^{-5}$) is indicated as a black line. Nearest genes are annotated for each locus, although the causal variants are not unequivocally known with the exception of *PNPLA3*. SNP, single nucleotide polymorphism.

Figure 2

Forest plots of odds ratios of the susceptibility loci for alcohol-related cirrhosis in comparison to alcohol misusers and population controls. Data generated using the rmeta package in R.

Figure 3

Fine-mapping analysis of the *MBOAT7* association signals. The –log10 of p values are plotted against the SNP genomic position based on NCBI build 37. The squares denote genotyped SNPs; the circles imputed SNPs (using 1,000 Genomes Project–based imputation). SNPs are coloured to reflect correlation with the most significant SNP with red denoting the highest LD ($r^2 > 0.8$) to the lead SNP. Estimated recombination rates from 1,000 Genomes Project (hg19/genomes Mar2012 EUR) are plotted in blue to reflect the local linkage disequilibrium structure. Gene annotations are obtained from the UCSC genome browser. Data generated using LocusZoom [38]. SNP, single nucleotide polymorphism; LD, linkage disequilibrium.

ONLINE METHODS

Phenotyping of cases and controls

Across all samples, cases were defined as patients with clinically diagnosed or biopsy-proven cirrhosis[48] on a background of past and/or present alcohol consumption of at least 60 g/day for women and 80 g/day for men for more than 10 years after exclusion of other causes of cirrhosis. Control individuals had no clinical or laboratory evidence of liver disease, confirmed by non-invasive assessment of liver fibrosis or examination of liver histology, on a background of alcohol dependence (per DSM-IV criteria) or reported alcohol consumption according to the criteria noted above. Case and control cohorts and details of recruitment protocols were reported previously for the German [9,23,41], UK [42] and Belgian [51] cohorts.

Germany, Austria and Switzerland

German/Austrian/Swiss patients comprised alcoholics with a long-term history of chronic alcohol abuse. Cases with alcoholic cirrhosis were recruited from gastroenterology and hepatology departments of the university hospitals in Dresden, Bonn, Leipzig, Kiel, Regensburg and Frankfurt (all Germany), Salzburg (Austria) and Bern (Switzerland). Control subjects drinking excessively, but without alcoholic liver cirrhosis, were recruited at psychiatry centres specialized in addiction medicine in Regensburg, Munich and Mannheim (all Germany), and Meiringen (Switzerland). Recruitment details were reported previously both for aspects related to liver disease [9] and those related to psychiatry [12]. Patients were recruited between 2000 and 2014 according to the same criteria across centres. Past and present alcohol consumption was quantified through interrogation during a face-to-face interview. All patients from Regensburg, Mannheim and Meiringen received a diagnosis of alcohol dependence (per DSM-IV criteria) by the consensus of 2 clinical psychiatrists. All

patients underwent careful clinical examination, standard laboratory testing and abdominal ultrasound. Chronic viral hepatitis was excluded in all patients by testing for hepatitis B surface antigen, anti-HBc and third-generation hepatitis C antibody ELISA. Serum levels of ferritin and transferrin saturation were determined to rule out hereditary hemochromatosis, and neither clinical nor serological signs of autoimmune liver disease were present. All patients and controls were of self-reported German ancestry.

Patients were defined as cases (alcoholic cirrhosis) when the presence of cirrhosis was assumed by at least one of the following criteria: 1. presence of cirrhosis as per liver biopsy (fibrosis stage 5 or 6 according to Ishak et al. [44]) or 2. unequivocal clinical and laboratory evidence for the presence of cirrhosis as reflected by a combination of (a) abnormal standard "liver" blood tests (transaminases, gamma-glutamyl-transpeptidase, coagulation tests, serum albumin concentration, platelet count), (b) cirrhosis-related complications including encephalopathy or ascites, (c) sonographic and/or radiological (computed tomography) findings suggestive of cirrhosis (hunched liver surface, ascites, splenomegaly), and (d) detection of esophageal varices via upper gastrointestinal endoscopy.

Corresponding control subjects were defined as alcoholic patients without cirrhosis when none of the criteria set forth for case patients were present. Patients gave written, informed consent and the study received approval from the ethics committees of all participating centres.

Belgium

Heavy drinkers with alcoholic cirrhosis (cases) and alcoholics without significant fibrosis (controls) with a history of excessive alcohol intake of ≥ 60 g/day (> 77 % drank more than 80 g/day) were recruited between 2002 and 2014 in Brussels and Haine-Saint-Paul (both Belgium). Cases were characterized as described above for the German/Austrian/Swiss

cohort. All heavy drinkers without clinically evident cirrhosis received a diagnosis of alcohol dependence based on DSM-IV criteria. They were screened with transient elastography (Fibroscan®) for the measure of liver stiffness (LSM) and/or liver biopsy to rule out significant liver fibrosis or cirrhosis related to excessive alcohol intake. Patients with a median LSM < 7 kPa were considered as having no or mild fibrosis [45]. In addition, 660 healthy Caucasians individuals from the general population were recruited at the same centres in Belgium. These individuals were recruited before minor surgery procedure or routine colonoscopy. They had neither clinical nor biochemical evidence (routine blood screening) of liver disease, nor evidence of other major pathological conditions. Written informed consent was obtained from all included subjects and the study received ethical approval.

United Kingdom

Subjects with self-reported English, Scottish, Welsh or Irish descent were recruited from the Centre for Hepatology at the Royal Free Hospital, London. Criteria applied for dissecting cases from control subjects were the same as outlined for the German/Austrian/Swiss cohort. The protocol was approved by the institutional review board and all included subjects consented to inclusion into the study.

Genotyping, quality control and GWAS

Genomic DNA was extracted from peripheral blood samples according to standard procedures and quantified using the PicoGreen dsDNA Quantitation Kit (Invitrogen Corporation, Carlsbad, California) and normalized to 50 ng/μL. Genotyping on Illumina bead chip arrays was performed according to the manufactures instructions and as reported before[23]. The first GWAS included 410 Germans with alcohol-related cirrhosis genotyped on the OmniExpress array (Version 12v1_j, Illumina Inc., San Diego, CA, USA) and 1,119 heavy drinkers without liver disease from a German psychiatric consortium genotyped on Illumina HumanHap550

BeadChip (N=407), Illumina Human610Quad (N=329) and Illumina Human660w Quad BeadChip arrays (N=383). In the second GWAS, 302 cases and 346 controls from the United Kingdom were genotyped using OmniExpress (version 24v1-0_a). To harmonize the German data sets, genotype probabilities were generated from signal intensity data from each array employing a Hidden Markov model of haplotype frequencies using BEAGLECALL[48] for all samples in a single computational batch. The genomic inflation factor λ calculated on the quality filtered dataset improved from 1.187 before to 1.1446 after BEAGLECALL for the German data set. We used EIGENSTRAT [49] to calculate principal components of genetic variation in the German genotype data. After correction for the first principal component the λ decreased 1.019 for the German case-control data set. Posterior genotype probabilities were converted to PLINK format hard calls using fcGENE [50] for initial quality control steps. Individuals with genotyping success < 97%, outlying autosomal heterozygosity (mean + 3*sd, mean-3*sd), kinship coefficient ($\hat{\pi}$) < 0.185 and those failing gender check were excluded from analysis. Samples that clustered outside the CEU HapMap population using multidimensional scaling (MDS) were likewise excluded. MDS analysis was performed on a cleaned LD pruned data set (indep-pairwise; excluding HLA region chr6:28477797-33448354; minor allele frequency >0.01, HWE $P > 1 \times 10^{-6}$, genotyping rate threshold for each marker > 95 %, genotyping rate threshold for each individual > 95 %) that was merged with Hapmap Phase III data from 11 different populations. Individuals deviating more than 3sd from the median European MDS cluster were excluded as population outliers. All QC filtering for the German and UK data sets was performed using PLINK (v1.07).

Imputation and GWAS metaanalysis in German and UK patients

To harmonize the four German GWAS data sets we selected only those SNPs that were genotyped on all arrays (N=298,405 quality controlled SNPs, intersection) as input for further imputation steps. Genotypes were imputed without pre-phasing to improve accuracy. Sample

phasing and genotype imputation was performed using IMPUTE 2 to reference 1000 Genomes Phase 3 (October 2014 release) or Phase 1 (March 2012 release) for X chromosome. For the German data set posterior genotype probabilities obtained by BEAGLECALL were used as input for IMPUTE v2.3.1. We controlled for between array differences among control samples genotyped on HumanHap550, Illumina Human610Quad and Illumina Human660w Quad by subsequently assigning alternating case control status to all samples on one array. We then performed case-control association analysis between arrays for all markers that are present on all three genotyping arrays to identify potential platform genotyping differences. All markers with $p < 0.001$ between all combination of control array pairs were identified as potential source of genotyping incongruities as recommended ("Method 3") by Sinnott and Kraft 2012.[46] These markers were excluded from further analysis. We also excluded markers with a HWE deviation of $P < 10-6$ in the combined control sample set.

Post-imputation quality filtering was performed using a minimum impute info score of 0.8, a Hardy-Weinberg equilibrium 1×10^{-6}, a minor allele frequency > 1%, yielding 6,866,424 SNPs for the German data set and 7,871,013 SNPs for the UK data set for further analysis. Quality controls and phenotypic association analysis on this marker set was performed with SNPTEST (v2.5) [47], using score statistic under an additive allelic effect model, including gender, age, BMI and type 2 diabetes status as covariates in the adjusted analyses.

Study-specific β - estimates from the German and UK scans were genomic-control adjusted prior to a fixed-effect model meta-analysis using an inverse variance weighted method implemented in META [22]. Meta-analysis was restricted to markers present in both data sets (N=6,770,425). The λ value was 1.005 for the combined meta-analysis.

At stage II the SNPs were validated in independent samples from Germany (N=1,290) and Belgium (N=779). Study-specific β – estimates and standard errors were derived from the German and Belgium stage II samples under an additive allelic effect model using PLINK

(v1.07) and further analysed using a fixed-effect model meta-analysis. Cochrane's Q and I square statistics were employed to assess consistency of effect and to quantify heterogeneity between both sample sets**.**

Validation Genotyping

One microliter of genomic DNA was amplified with the GenomiPhi (Amersham, Uppsala, Sweden) whole genome amplification kit and fragmented at 99°C for three minutes. Genotyping of SNPs rs738409 (hcv7241), rs4823173 (hcv25931728), rs10401969 (hcv30444981), rs58542926 (hcv89463510), rs6556045 (hcv29201253), rs6605237 (hcv27032502), rs17886348 (hcv25618641), rs7769670 (hcv29432279), rs7845021 (hcv1383849), rs626283 (hcv2916337), rs8736 (hcv2916339), rs641738 (hcv8716820), rs62190923 (Assay by Design), rs7812374(AdD), rs34564463(AdD), rs739846(AdD(failed)) was performed using the Taqman chemistry (Applied Biosystems, Foster City, Ca, USA) on an automated platform. Reactions were completed and read in a 7900 HT TaqMan sequence detector system (Applied Biosystems). The amplification reaction was carried out with the TaqMan universal master mix. Thermal cycling conditions consisted of 1 cycle for 10 minutes at 95°C, 45 cycles for 15 seconds at 95°C, and 45 cycles for 1 minute at 60°C. Genotyping of SNPs rs709186, rs9874300, rs5867471, rs71313753, rs2531812, rs16849557, rs1177809, rs1560408, rs6846459, rs36640, rs254283, rs8112480, rs58694079, rs4806498, rs10416555, rs1050527, rs626283, rs36654, rs36633, rs79199039, rs2576452, rs73062690 was performed using the Sequenom iPLEX Gold chemistry MassARRAY platform. All process data were logged and administered with a database-driven LIMS [18].

Software and statistical analysis

BEAGLECALL (1.0.1) was used for genotype calling of the German data set. SNP imputation was performed with IMPUTE v2.3.1. QC filtering was performed using PLINK

(1.07) and on BEAGLE and SNPTEST output data. Processing and transformation of SNP data was accomplished by using fcGENE (1.0.7). Stage I association tests were performed by SNPTEST (2.5), stage II logistic regression tests were performed using PLINK (1.07). The program META (1.6.0) was used for meta-analysis of stage I, II and combined data sets.

Forest plots were generated using rmeta package in R. Regional fine mapping plots were generated using LocusZoom (1.1)[38]. Population attributable risk (PAR) was calculated as $PAR=P_{RF}\times(RR-1)/1+P_{RF}\times(RR-1)$. The Population Attributable Risk percentage (PAR%) was calculated as $PAR=P_{RF}\times(RR-1)/1+P_{RF}\times(RR-1)\times100\%$ where P_{RF} is the population prevalence of the risk factor in alcohol misusers or in the general population for all three analyzed populations, respectively. The allelic odds ratio was used as an approximation for the estimate relative risk (RR) of disease due to exposure of the risk allele. Combined PAR was calculated by the formula $PAR_{combined} = 1-(1-PAR_1)\times(1-PAR_2)\times(1-PAR_3)$.

Expression analysis of MBOAT7 and TMC4 transcripts

The qRT was performed using specific primer for TMC4 and MBOAT7 and normalized to ACTB in cDNA from patients with alcoholic liver cirrhosis bearing the homozygous ancient, heterozygous and homozygous mutant genotype for rs641738 (CC n=5; CT n=6, TT n=6). Primer sequences are provided in *Supplementary Table 7*.